(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/37465 A2

(51) International Patent Classification[7]: H04H

(21) International Application Number: PCT/EP00/10617

(22) International Filing Date: 26 October 2000 (26.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/442,960 18 November 1999 (18.11.1999) US

(71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventors: DIMITROVA, Nevenka; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). MCGEE, Thomas; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). ELENBAAS, Jan, H.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). AGNIHOTRI, Lalitha; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). JASINSCHI, Radu; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). DAGTAS, Serhan; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). MENDELSOHN, Aaron; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: SCHMITZ, Herman, J., R.; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States *(national)*: CN, JP, KR.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/37465 A2

(54) Title: METHOD AND APPARATUS FOR AUDIO/DATA/VISUAL INFORMATION SELECTION

(57) Abstract: A method of selecting, storing and delivering desired audio/data/visual information includes the steps of determining viewing preferences of a viewer (100) and receiving a first group of audio/data/visual signals (102), for example, broadcast and cable television signals or internet-based signals. Based on the first group of audio/data/visual signals, a second group of audio/data/visual signals, which is a subset of the first group of audio/data/visual signals, is identified (108). The second group of audio/data/visual signals is selected based on the association of EPG data for each signal with the viewing preferences of the viewer. Content data is then extracted from the second group of audio/data/visual signals and compared with the viewing preferences (110, 114). The content data may include, for example, closed-captioned text, EPG data, audio information, visual information and transcript information. Based on the comparison of the content data extracted from the second group of audio/data/visual signals with the viewing preferences, audio/data/visual information contained in the second group of audio/data/visual signals which is of interest to the viewer is identified (122) and stored for review at the viewers convenience (124).

Method and apparatus for audio/data/visual information selection

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates generally to an audio/data/visual information selection system, and more particularly to a system which filters a stream of audio/data/visual signals, such as television, data or internet-based signals, and provides filtered information, based on user defined parameters, at various levels of content abstraction.

Background Information

As the number of television channels increases, a television viewer has a dilemma as to what television shows to watch and how to make the best use of his time while watching television. Since printed and on-screen television listings do not accurately identify the entire content of each television program, many television viewers have taken to "channel surfing" to identify television programs or portions thereof that are "of interest".

Oftentimes, a television viewer spends a great amount of time channel surfing in the hope of identifying television programs which correspond to his interest. This may cause the viewer to miss many other television programs that he would enjoy watching. For example, if there are commercials that are airing on the "surfed" channels, the viewer will encounter delays in identifying the television program that is being broadcast on the "surfed" channel. Therefore it takes longer to determine whether the program being broadcast is of interest. As a result, programming which may be of interest that is broadcast on other channels will be missed. If the viewer does locate a desirable television program, he often encounters uninteresting commercials that are aired during the show, thereby missing a concurrently aired program of interest that is being broadcast on another channel.

Television viewers are generally tired of the ever-increasing number of television channels that have programming of interest only a portion of the time, the multitude of commercials that are aired during programming, and channel surfing. Therefore, a technique for the scanning, smart selection and/or recording of broadcast television and cable programs and/or information that are of interest to a viewer is essential for the television of the future.

Although there have been recent improvements in digital video processing as is evident by new capture boards and fast processors, relatively little advancement has been made on how the information conveyed by video data can best be recovered, analyzed, classified and delivered according to a viewer's desires.

Systems have recently been developed wherein electronic program guide (EPG) data is analyzed based on viewer information that is provided to the system. Based on the analysis of the EPG data, a list of television programs which may be of interest to the viewer is provided. The EPG data is, however, limited and does not enable different levels of content analysis of every video frame or segment of each television program based on viewer defined parameters.

OBJECTS AND SUMMARY OF THE INVENTION

It is an object of the present invention to provide a method and apparatus for audio/data/visual information selection, storage and delivery that overcomes the aforementioned problems with the prior art.

It is another object of the present invention to provide a method and apparatus for audio/data/visual information selection, storage and delivery that monitors a plurality of audio/data/visual signals, identifies audio/data/visual information that is of interest to an individual, and enables use of the identified audio/data/visual information by the individual.

It is another object of the present invention to provide a method and apparatus that selectively records only segments of television-based and/or internet-based information that correspond to the defined parameters.

In accordance with one form of the present invention, a method of selecting desired audio/data/visual information that are of interest and that reflect personal preferences and taste in terms of television programs includes the steps of determining viewing preferences of a viewer, receiving a first plurality of audio/data/visual signals, identifying from the first plurality of audio/data/visual signals a second plurality of audio/data/visual signals to be monitored wherein the second plurality of audio/data/visual signals is a subset of the first plurality of audio/data/visual signals, comparing the viewing preferences with the second plurality of audio/data/visual signals to identify desired audio/data/visual information, and providing access to the desired audio/data/visual information.

In accordance with another aspect of the present invention, a method of selecting desired audio/data/visual information includes the steps of determining preferences of a user, receiving a plurality of audio/data/visual signals, comparing the preferences with

the plurality of audio/data/visual signals to identify desired audio/data/visual information, and providing access to the desired audio/data/visual information.

In accordance with another aspect of the present invention, an audio/data/visual signal selection system includes an input device for providing viewing preferences of a viewer, and an information selector. The information selector receives a first plurality of audio/data/visual signals, identifies from the first plurality of audio/data/visual signals a second plurality of audio/data/visual signals to be monitored wherein the second plurality of audio/data/visual signals is a subset of the first plurality of audio/data/visual signals. The information selector also compares the viewing preferences with the second plurality of audio/data/visual signals to identify desired audio/data/visual information, and provides access to the identified audio/data/visual information.

In accordance with another aspect of the present invention, an audio/data/visual signal selection system includes an input device for providing preferences of a user, and an information selector. The information selector receives a plurality of audio/data/visual signals, compares the preferences and the plurality of audio/data/visual signals to identify desired audio/data/visual information, and provides access to the identified audio/data/visual information.

The above and other objects, features and advantages of the present invention will become readily apparent from the following detailed description thereof, which is to be read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram of the audio/data/visual information selection system according to the present invention; and

Fig. 2 is a flow chart of the operation of the audio/data/visual information selection system of Fig. 1.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention provides an audio/data/visual information selection system ("video scouting system") which monitors a plurality of television or internet-based audio/data/visual signals. The system selects and records audio/data/visual information which may be of interest to a viewer based on the preferences of the viewer. For example, when a person desires to watch television programs relating to certain topics, the person need not know the television programs, the show times and/or the television channels. Moreover,

the person need not be present at the time the programs are broadcast. The system of the present invention sets up a programmable wish list for the programs, personalities, characteristics, topics, and events that the person desires to watch. The system then continually monitors the received television signals for items on the wish list, records the entire television program or portions thereof which meet the criteria of the wish list, and enables access to the recorded items for viewing by the viewer at a convenient time.

The audio/data/visual information selection system according to the present invention may be contained within a computer or television, or it may be a stand-alone device coupled to the television or computer, that "surfs" received television, radio or internet-based signals and records desired segments of programs on a local storage device. The selection of program segments is based on content data of the broadcast and cable television or internet-based signal. The content data may include closed-captioned text, EPG data (which can be in the form of meta-data), audio information (such as frequency, pitch, timbre, sound, and melody), visual information (such as color, motion, shape, and texture of 2-D/3-D objects) and transcript information.

While the present invention is described herein in the context of its use in connection with television broadcast signals, it is foreseen that the system can be utilized with computers that have internet accessibility so as to scan internet-based signals (for example as a web crawler or video web portal) for information that is of interest to the user, radios for personalized radio applications so as to scan for particular types of audio signals, information networks (e.g. proprietary networks and personal area networks), and for systems which only transmit data information.

Referring now to Fig. 1 of the drawings, the apparatus for audio/data/visual information selection 10 is shown. The apparatus is preferably coupled to a video monitor 12, such as a television. The apparatus includes a personal profile input device 14 (for example a keypad, keyboard, on-screen display, television remote control, touchscreen, verbal command receiver or touchpad) by which a user can input personal viewing preferences. The viewing preferences correspond to characteristics of television programs that the user desires to watch (and/or has watched before). A few examples of viewing preferences include sports teams (e.g. New York Yankees), individuals (e.g., President Bill Clinton), locations (e.g., the White House), actions (e.g., a player hitting a home run), and topic (e.g., World Championship). Based on all of the data provided by the user, a profile of the user's viewing preferences is generated. As explained in detail below, the received television signals are monitored for, at least, segments of broadcasts which correspond to the

viewing preferences. For example if a viewing preference was "New York Yankees", the present invention would record an entire New York Yankee baseball game, without commercials unless the commercial relates to the New York Yankees, any movie involving the New York Yankees (e.g., "Pride of the Yankees") and the sports portion of a news/sports show which shows highlights of the previous night's New York Yankee baseball game.

It is also foreseen that the apparatus has personal profiles stored in memory for a variety of topics so the user need not input specific viewing preferences, but need only input a number corresponding to a predetermined type of viewing preference (for example, a "1" for sports, "2" for local news, "3" for national news, "4" for world news, "5" for politics, "6" for science fiction, etc.). As a result, the user need not enter a great amount of information for the system to provide a broad range of desired information.

The apparatus also includes an audio/data/visual signal receiver 16 for receiving a plurality of television signals to be analyzed. Instead of filtering out all but one of the signals as performed by a television, the receiver 16 has multiple tuners and maintains all signals for analysis. Suitable audio/data/visual signal receivers include an antenna, satellite dish, set-top box, internet connection, cable and the like. As known in the art, the broadcast and cable television signals provided to the receiver are multiplexed signals.

Operatively coupled to the output of the audio/data/visual signal receiver 16 is a demultiplexer 18 for demultiplexing the multiplexed plurality of television signals received by the audio/data/visual signal receiver. The demultiplexer demultiplexes the plurality of signals and enables each of the plurality of television signals to be individually analyzed as explained in detail below.

In the preferred embodiment the apparatus includes an EPG (electronic programming guide) signal receiver 20 for receiving electronic programming guide signals associated with the plurality of television signals. As known in the art, the EPG signals include a vast assortment of information about the television programs currently being aired and that are to be aired. Examples of EPG information include the title, start time, end time, actors (if applicable), topic, category of program and a brief program description. Suitable EPG signal receivers include an antenna, satellite dish, set-top box, internet connection and the like. It is foreseen that the EPG signal receiver and audio/data/visual signal receiver could be combined into one device wherein the combined device switches between a first mode for receiving audio/data/visual signals and a second mode for receiving EPG signals. Alternatively, the device could concurrently receive audio/data/visual signals and EPG signals.

The apparatus also includes an audio/data/visual information selector 22 which receives the EPG signals from the EPG signal receiver 20, the demultiplexed television signals from the demultiplexer 18 and the viewer preferences from the personal profile input device 14. The audio/data/visual information selector analyzes the demultiplexed audio/data/visual signals based on their content data (explained in detail below), the corresponding EPG signals and the viewing preferences to identify television broadcasts which are of interest to the user. The information might be an entire television program or it could only be a segment thereof if the EPG data indicates that only a segment of the television program corresponds to the viewing preferences. It is important to note that the audio/data/visual information selector is preferably capable of concurrently analyzing each of the television signals provided thereto so as to monitor the television signals in parallel, and to record in a memory the television information identified by the audio/data/visual information selector. An example of a suitable device that can be programmed to perform the functions of the audio/data/visual information selector is a CPU (for example, Pentium or MIPS) of a personal computer, a special programmable Digital Signal Processor (such as Trimedia) or a specially configured chip architecture. The operation of the audio/data/visual information selector will be explained in the detail later.

Operatively coupled to the audio/data/visual information selector 22 is a memory 24 (for example, RAM, hard disk recorder, optical storage device, or DVHS, each having hundreds of giga bytes of storage capability) for recording the television broadcasts or portions thereof identified by the audio/data/visual information selector 22 as corresponding to the viewing preferences. When requested by the user, the audio/data/visual information selector can access the audio/data/visual information stored in the memory and provide the information to the video monitor 12 for review by the user.

Referring now to Fig. 2, the operation of the apparatus for audio/data/visual information selection, storage and delivery will be described.

Initially, a user inputs personal profile data via the personal profile input device 14 (Step 100). The personal profile corresponds to the viewing preferences of the user such as specific types of television programs, persons or aspects of televisions programs that the viewer desires to watch. This information can be provided in numerous ways. For example, the information can be input via personal profile input device 14. Alternatively, the information can be input through an onscreen guide on the television or video monitor 12 via arrow keys on a conventional television remote control device. Alternative to the above, all of the user profile information can be automatically generated wherein the personal profile

input device monitors the viewing habits of the user and, through artificial intelligence, "learns" the personal viewing preferences of the user. It is foreseen that the user profile can evolve based on user behavior and changing viewing interests. It is also foreseen that the information selector or input device monitors the user's changing viewing habits and automatically updates the viewing preferences (Step 101). For example, if the user previously watched only sporting events but has recently been watching a business news channel, the system will modify the original viewing preference (sporting events) to include business news. The personal profile input device preferably stores the "learned" viewing habits in an internal memory (not shown). Alternatively, the "learned" viewing habits are stored in memory 24. It is also foreseen that the user can select one of the pre-stored profiles based on his closest match.

The audio/data/visual signal receiver 16 receives the audio/data/visual television signals available to the viewer (for example, those channels that the viewer subscribes to through the local cable television or satellite service) (Step 102), the television signals are demultiplexed by demultiplexer 18 (Step 104) and provided to the audio/data/visual information selector 22 (Step 106). The EPG signals are received by EPG signal receiver 20 which are then provided to the audio/data/visual information selector 22 (Step 106).

The audio/data/visual information selector 22 performs an initial selection process of the received television signals. Relevant portions of the EPG data for each of the received television signals are compared with the viewing preferences to determine which television programs are not at all associated with the viewing preferences (Step 108). The television programs which are not associated with the viewing preferences are not monitored. As a result, a subset of the received television signals (television programs) is maintained for further analysis and conformance with the viewing parameters. The use of the EPG data is only a first level of filtering to eliminate television programs which are clearly not at all associated with the viewing preferences of the user. For example, if the EPG data for one television signal identifies the corresponding television channel as airing the movie "Gone With the Wind", and the viewing preferences of the user are related to "baseball" or the "stock market", there is no need to monitor this channel while the movie is being shown. However, if the EPG data for another television signal identifies the corresponding channel as currently broadcasting the news, monitoring of this channel would be warranted since the previous night's baseball scores and the day's business news may be discussed.

It should be noted that if EPG data is not available to determine the subset of received television signals to be analyzed, then the audio/data/visual information selector initially monitors a group of preferred channels identified in the viewing preferences or the channels frequently watched by the user. If there are no limits on computation resources of the audio/data/visual information selector, then all available channels will be concurrently monitored.

It should also be mentioned that the non-monitored television programs are periodically checked (i.e., reviewed) to ensure that the programming on the corresponding channel has not changed and is not now broadcasting a program which coincides with the viewing preferences.

Once a subset of television channels has been selected, each of the subset of television channels is continually analyzed in parallel to determine which (if any) portions of the currently aired program correspond to the viewing preferences (each television program is concurrently analyzed). The analysis includes extracting closed-captioned or transcribed text from each television program to be analyzed (Step 110). The extracted closed-captioned or transcribed text is indexed (Step 112). Specifically, indexing, as known in the art, includes monitoring the frequency of occurrence of words in the text so as to provide an indication of the subject matter of the program. Indexing is explained in the publications entitled "Introduction to Modern Information Retrieval" by G. Salton and M.J. McGill, McGraw-Hill, NY, NY, 1983; "Natural Language Understanding" by James Allen, The Benjamin/Cummings Publishing Company, Inc., 1995; and "Advances in Automatic Text Summarization", edited by Inderjeet Mani and Mark T. Maybury, The MIT Press, Cambridge, MA, 1999, the entire disclosures of which are incorporated herein by reference. The indexed text is analyzed to determine whether particular words are frequently used in the programs which have an association with the viewing preferences (Step 114). If frequently used words in the television program do coincide with the viewing preferences, then the program or relevant segment should be noted and either further analyzed or recorded.

Concurrent to the text extraction and indexing, the television programs are monitored for the occurrence of commercials (Step 116). If the viewing preferences do not include an interest in commercials, when a commercial is aired on one of the television channels being analyzed, the present invention does not analyze the commercials so that system resources can be concentrated on the noncommercial television broadcast. Otherwise if the commercials are desired, all commercials can be stored in memory for analysis at a later time.

The method also includes segmentation of the video portion of the television signal (Step 118) to analyze video frames of the television program. In the preferred embodiment, every video frame of each program being monitored is analyzed (that is, in the U.S., 30 video frames are analyzed per second). Video segmentation is known in the art and is generally explained in the publications entitled, "Parsing TV Programs For Identification and Removal of Non-Story Segments", by T. McGee and N. Dimitrova, Proc. of SPIE Conf. on Storage and Retrieval for Image and Video Databases, pp. 243-251, San Jose, CA, January, 1999; "PNRS-Personal News Retrieval System", by N. Dimitrova, H. Elenbaas and T. McGee, SPIE Conference on Multimedia Storage and Archiving Systems IV, pp. 2-10, September 1999, Boston; and "Text, Speech, and Vision For Video Segmentation: The Infomedia Project" by A. Hauptmann and M. Smith, AAAI Fall 1995 Symposium on Computational Models for Integrating Language and Vision 1995, the entire disclosures of which are incorporated herein by reference. If the user's viewing preferences indicate a desire to view subject matter on John F. Kennedy, any segment of the video portion of the television program including visual (e.g., a face) and/or text information relating to John F. Kennedy will indicate that the current broadcast relates to the user's viewing preferences. As known in the art, video segmentation includes, but is not limited to:

Cut detection: wherein two consecutive video frames are compared to identify abrupt scene changes (hard cuts) or soft transitions (dissolve, fade-in and fade-out). An explanation of cut detection is provided in the publication by N. Dimitrova, T. McGee, H. Elenbaas, entitled "Video Keyframe Extraction and Filtering: A Keyframe is Not a Keyframe to Everyone", Proc. ACM Conf. on Knowledge and Information Management, pp. 113-120, 1997, the entire disclosure of which is incorporated herein by reference.

Face detection: wherein regions of the video frames are identified which contain skin-tone and which correspond to oval-like shapes. In the preferred embodiment, once a face image is identified, the image is compared to a database of known facial images stored in the memory to determine whether the facial image shown in the video frame corresponds to the user's viewing preference. An explanation of face detection is provided in the publication by Gang Wei and Ishwar K. Sethi, entitled "Face Detection for Image Annotation", Pattern Recognition Letters, Vol. 20, No. 11, November 1999, the entire disclosure of which is incorporated herein by reference.

Text detection: wherein text which appears in the video frame such as overlayed or superimposed text is identified and a determination is made as to whether the text is related to the user's viewing preferences. An explanation of text detection is provided

in the article entitled "Text Detection in Video Segments" by L. Agnihotri and N. Dimitrova, Proceedings of IEEE Workshop on CBAIVL, Fort Collins, Colorado, June 1999, held in conjunction with IEEE Conference on Computer Vision and Pattern Recognition 1999, the entire disclosure of which is incorporated herein by reference. In the preferred embodiment, once the text is detected, optical character recognition (OCR) which is known in the art is employed on the detected regions and a look-up table stored in memory is used to identify the detected text. The look-up table preferably includes associations between a variety of words. For example, "Bill Clinton" may be associated with "President of the United States" and "politics", "White House", "Monica Lewinsky" and "Whitewater".

Motion Estimation/Segmentation/Detection: wherein moving objects are determined in video sequences and the trajectory of the moving object is analyzed. In order to determine the movement of objects in video sequences, known operations such as optical flow estimation, motion compensation and motion segmentation are preferably employed. An explanation of motion estimation/segmentation/detection is provided in the publication by Patrick Bouthemy and Francois Edouard, entitled "Motion Segmentation and Qualitative Dynamic Scene Analysis from an Image Sequence", International Journal of Computer Vision, Vol. 10, No. 2, pp. 157-182, April 1993, the entire disclosure of which is incorporated herein by reference.

Camera Motion: wherein a set of five (5) global camera parameters are employed, preferably two (2) translational and three (3) rotational. The 3-D camera motion is then classified as pure tracking (horizontal motion), booming (vertical motion), dollying (motion in depth), panning (rotation about the vertical global axis), tilting (rotation about the horizontal axis), and rolling (rotation about the z-axis) or combinations of these motions. This information can be used to classify the video shots into, for example, "static", "zoom" and/or "span", and to further determine the director's intention for producing the shot. The camera motion information is used in classification such that if EPG data is not available, the category of the program can be determined based on camera motion. An explanation of camera motion detection is provided in the publication by R. Y. Tsai and T.S. Huang entitled "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces", IEEE Transaction on PAMI, Vol. 6, number 1, pp. 13-27, 1994, the entire disclosure of which is incorporated herein by reference.

The method also includes segmentation of the audio portion of the television signal (Step 120) wherein the audio portion of the television broadcast is monitored for the occurrence of words/sounds that are relevant to the viewing preferences. Audio

segmentation includes the following types of analysis of television programs: speech-to-text conversion, audio effects and event detection, speaker identification, program identification, music classification, and dialog detection based on speaker identification.

Audio segmentation includes division of the audio signal into speech and non-speech portions. The first step in audio segmentation involves segment classification using low-level audio features such as bandwidth, energy and pitch. Thereafter channel separation is employed to separate simultaneously occurring audio components from each other (such as music and speech) such that each can be independently analyzed. Thereafter, the audio portion of the television program is processed in different ways such as speech-to-text conversion, audio effects and events detection, and speaker identification. Audio segmentation is known in the art and is generally explained in the publication by E. Wold and T. Blum entitled "Content-Based Classification, Search, and Retrieval of Audio", IEEE Multimedia, pp. 27-36, Fall 1996, the entire disclosure of which is incorporated herein by reference.

Speech-to-text conversion (known in the art, see for example, the publication by P. Beyerlein, X. Aubert, R. Haeb-Umbach, D. Klakow, M. Ulrich, A. Wendemuth and P. Wilcox, entitled "Automaic Transcription of English Broadcast News", DARPA Broadcast News Transcription and Understanding Workshop, VA, Feb. 8-11, 1998, the entire disclosure of which is incorporated herein by reference) can be employed once the speech segments of the audio portion of the television signal are identified or isolated from background noise or music. Speech-to-text conversion is important if closed-captioning is not available to provide a transcript of the audio portion of the television program. The speech-to-text conversion can be used for applications such as keyword spotting with respect to the viewing preferences.

Audio effects can be used for detecting events (known in the art, see for example the publication by T. Blum, D. Keislar, J. Wheaton, and E. Wold, entitled "Audio Databases with Content-Based Retrieval", Intelligent Multimedia Information Retrieval, AAAI Press, Menlo Park, California, pp. 113-135, 1997, the entire disclosure of which is incorporated herein by reference). Events can be detected by identifying the sounds that may be associated with specific events. For example, an announcer shouting "goal" in a sporting event could be detected and the program segment could then be recorded in memory if the viewing parameters include replays of hockey or soccer goals.

Speaker identification (known in the art, see for example, the publication by Nilesh V. Patel and Ishwar K. Sethi, entitled "Video Classification Using Speaker Identification", IS&T SPIE Proceedings: Storage and Retrieval for Image and Video

Databases V, pp. 218-225, San Jose, CA, February 1997, the entire disclosure of which is incorporated herein by reference) involves analyzing the voice signature of speech present in the audio signal to determine the identity of the person speaking. Speaker identification can be used, for example, to search for a favorite actor or the comments of a political figure.

Program identification involves analyzing the audio portion of the audio/data/visual signal to identify a television program. This is especially useful in cataloging and indexing of programs. This is important if EPG information is not available. The analyzed audio portion is compared to a library of program characteristics to identify the program to determine if the program coincides with the viewing parameters.

Music classification involves analyzing the non-speech portion of the audio signal to determine the type of music (classical, rock, jazz, etc.) present. This is accomplished by analyzing, for example, the frequency, pitch, timbre, sound and melody of the non-speech portion of the audio signal and comparing the results of the analysis with known characteristics of specific types of music. Music classification is known in the art and explained generally in the publication entitled "Towards Music Understanding Without Separation: Segmenting Music With Correlogram Comodulation" by Eric D. Scheirer, 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY October 17-20, 1999.

After segmentation of the audio and video signals, various portions of the segmented audio and video signals are combined (integrated) (Step 121), if appropriate, to determine if the current television program coincides with the viewing parameters. Integration of the segmented audio and video signals is necessary for complex viewing parameters. For example, if the viewer desires to see a particular actor speak a particular line while making a particular hand gesture, not only is face recognition required (to identify the actor) but also speaker identification (to ensure the actor on the screen is speaking), speech to text conversion (to ensure the actor speaks the appropriate words) and motion estimation/segmentation/detection (to recognize the specified hand gesture of the actor).

As explained above, concurrent to the visual and audio segmentation, the segmented audio and video information is integrated (if applicable) and compared to the viewing preferences (Step 122). Thereafter a determination is made as to whether the current program on the particular channel coincides with the viewing preferences. If so, in the preferred embodiment the program is ranked for its conformance with the viewing preferences (Step 123) and is recorded in memory for as long as the current program coincides with the user's viewing preferences (Step 124). The ranking involves placing

video segments which correspond to the user's most favored viewing parameters first in the memory and those related to the least favored viewing preferences last. In this way, when the viewer watches the recorded program segments, the highest ranked will be viewed first.

If the user is concurrently watching another television show while the present invention is operating, the user could concurrently be informed that a television show of interest is being broadcast on a specific channel (Step 126). Upon request by the viewer, the recorded program segments are retrieved and provided for viewing by the user (Step 128).

As mentioned above, it is foreseen that portions of the audio and video segmentation (determined by core AV modules) can be combined (that is, integrated) using mid-level and high-level modules to determine specific events during the television program. The core AV modules include visual and textual modules that provide independent audio and video analysis. These modules by themselves include multiple processing units. The main purpose of the core AV modules is to extract lower-level features that can be used as input to integration modules (mid-level and high-level tools) for inferring higher-level decisions that resemble semantic descriptions of the television program content. The basic visual attributes are color, motion, shape, and texture. Each of these attributes is described by a large set of operators that range from local operators to regional/global operators. These operators are primitive because they are processed independently from each other and they are task independent. The set of mid-level and high-level integration modules contain tools that combine different elements from the core AV modules. The purpose of the integration modules is to extract high-level information from the content data. This involves multimodal integration of lower-level features. The mid-level tools (modules) typically are used to describe relationships between object parts and audio/data/visual attributes. The high-level tools are used to identify/relate/process objects. These models can be static or dynamic. The dynamic models are updated in time.

What distinguishes high-level from mid-level information is that for the former there exists a decision process in the loop. This means that, internal to the module, there exists a process of deciding which core AV modules to use and under what conditions. Typical examples of the high-level modules are action, event detection/recognition, story segmentation and classification, program classification, and context detection.

If a viewing preference is a specific action of an actor (e.g., opening a door and entering a room), mid-level or high-level modules would be used because not only are face detection and/or voice identification employed, but motion detection is used to pinpoint the action of the particular actor to determine if the action of the actor corresponds with the

viewing parameter. As a result, multiple decision loops are employed to analyze the television program.

It is foreseen that the present invention is capable of developing video summaries of entire programs so that the recorded segments that are viewed by the user can be watched in the context of the program (i.e., a "catch-up" function). The video summaries can be developed using key frame images and closed-captioned text to provide an indication of the portions of the program that were not recorded.

It is also foreseen that the viewing preferences can be automatically updated each time that a particular user watches television. This is accomplished based on the viewer's time spent watching certain programs (and the category of programs) as well as the visual and other characteristics of the program (for example, action, bright colors). Parental control can be added to filter out parts of television programs or entire programs based on the content of the program. The present invention can detect scenes of television programs that have nudity, violence or obscene words and prevent those portions of the program from being viewed by minors.

It is foreseen that the system can provide updates to the viewer regarding the information that was recorded while the viewer is watching television. In other words, the viewer is informed that television segments have been recorded which match the viewing parameters while the viewer is watching another television channel. It is also foreseen that if a user is watching one television program and the system identifies a program of interest, the user is informed in real-time of the detection of the program of interest. Further, it is foreseen that the system performs a daily/weekly automatic storage cleanup function of the memory to manage storage space based on the viewing preferences. In the preferred embodiment the system also includes a time catch-up function. Specifically, when the person is surfing television channels and stumbles upon an interesting program, the person can "catch up" by viewing "an extracted video poster" (or abstract, trailer).

The present invention therefore provides a "video scouting system" wherein when a person desires to watch certain types of television programs or only wants to view specific information, the present invention sets up a programmable wish list for the programs, topics and events that the viewer desires watching.

Although the present invention has been discussed relative to finding desirable television programs and television program segments/information for a viewer at the viewer's location, a video brokerage house service could be used for filtering and delivery of specific video segments on demand. Therefore the system would not be located at the user's end, but

at, for example, the cable television providers end and the system would operate concurrently for a plurality of users.

The present invention uses and adapts existing techniques such as video segmentation, video parsing, speech recognition, character recognition, and object spotting, for finding cues in the video streams to provide a personalized video information identification system.

Further, it is foreseen that the present invention can be adapted to monitor and record the viewer's feedback and interest in specific programs. The collected information can be valuable for launching new programs, new products, new films and the production of specific events. The present invention is able to capture individual viewer interests. The more the apparatus is used by a person, the better able it is to adapt to the user's diverse interests. The present invention thus models the information about the individual interests that change over time, both as users change and as the system acquires more information about users. This information is useful to advertisers and broadcasting companies. The video filtering system is able to produce a browsable layout of an entire movie or a television program by a video analysis process. Further, the present invention permits a person to preview and prehear the content of a television program as a multimedia presentation. This is achieved by segmenting the video, analyzing its contents and presenting the user with a browsable layout consisting of original and synthetic frames, as well as important conversation segments.

Having described specific preferred embodiments of the invention with reference to the accompanying drawings, it will be appreciated that the present invention is not limited to those precise embodiments and that various changes and modifications can be effected therein by one of ordinary skill in the art without departing from the scope or spirit of the invention defined by the appended claims.

CLAIMS:

1. A method of selecting desired audio/data/visual information comprising the steps of:

a) determining viewing preferences of a viewer (100);

b) receiving a first plurality of audio/data/visual signals (102);

c) identifying, from the first plurality of audio/data/visual signals, a second plurality of audio/data/visual signals to be monitored, wherein the second plurality of audio/data/visual signals is a subset of the first plurality of audio/data/visual signals (108);

d) comparing said viewing preferences with the second plurality of audio/data/visual signals to identify desired audio/data/visual information (122); and

e) providing access to the desired audio/data/visual information (124).

2. The method of selecting desired audio/data/visual information according to claim 1 further comprising the steps of:

storing the desired audio/data/visual information (124); and

retrieving the desired audio/data/visual information when requested by the viewer (128).

3. The method of selecting desired audio/data/visual information according to claim 1 wherein step (d) comprises the steps of:

extracting content data corresponding to each of said second plurality of audio/data/visual signals (110);

indexing said extracted content data (112); and

comparing said indexed content data with said viewing preferences to identify desired audio/data/visual information (114).

4. The method of selecting desired audio/data/visual information according to claim 3 wherein the extracted content data comprises at least one of closed-captioned text, EPG data, audio content information, visual content information and transcript information.

5. The method of selecting desired audio/data/visual information according to claim 4 wherein the visual content information comprises at least one of cut detection, face detection, text detection, motion estimation/segmentation/detection and camera motion.

6. The method of selecting desired audio/data/visual information according to claim 4 wherein the audio content information comprises at least one of speech-to-text conversion, audio effects and event detection, speaker identification, program identification, music classification and dialog detection based on speaker identification.

7. The method of selecting desired audio/data/visual information according to claim 4 wherein the transcript information comprises at least one of natural language processing and understanding, discourse analysis, keyword detection, and broadcast categorization.

8. The method of selecting desired audio/data/visual information according to claim 4 further comprising the step of:

integrating at least two of the closed-captioning text, the EPG data, the extracted audio information, the extracted visual information, and the extracted transcript information (121).

9. The method of selecting desired audio/data/visual information according to claim 8 wherein the integrating step provides at least one of event and action detection, story segmentation, story classification, program classification, and context detection.

10. The method of selecting desired audio/data/visual information according to claim 9 wherein the context detection comprises detection of at least one of human faces and scenery.

11. The method of selecting desired audio/data/visual information according to claim 1 wherein said desired audio/data/visual information comprises at least one of broadcast and cable television signals, internet-based signals and data signals.

12. The method of selecting desired audio/data/visual information according to claim 1 wherein step (c) comprises the step of:

comparing electronic programming guide (EPG) data for each of the first plurality of audio/data/visual signals with the viewing preferences to identify the second plurality of audio/data/visual signals which are associated with the viewing preferences (108).

13. The method of selecting desired audio/data/visual information according to claim 1 wherein step (a) comprises the step of:

inputting, via a keypad, keyboard, on-screen display, remote control, touchscreen, verbal commands or touchpad, characteristics of audio/data/visual information that the viewer desires to watch (100).

14. The method of selecting desired audio/data/visual information according to claim 1 wherein step (a) comprises the step of:

monitoring the viewing habits of the viewer to formulate the viewing preferences which correspond to characteristics of audio/data/visual information that the viewer desires to watch (101).

15. The method of selecting desired audio/data/visual information according to claim 14 further comprising the step of:

automatically updating the viewing preferences each time that the viewer accesses television broadcast signals or internet-based signals (101).

16. The method of selecting desired audio/data/visual information according to claim 1 further comprising the step of:

identifying commercial and noncommercial portions of the second plurality of audio/data/visual signals (116).

17. The method of selecting desired audio/data/visual information according to claim 1 further comprising the step of:

ranking the desired audio/data/visual information according to relevance thereof to said viewing preferences (123).

18. The method of selecting desired audio/data/visual information according to claim 1 wherein step (e) comprises the step of:

storing at least a portion of said desired audio/data/visual information in a memory (124).

19. The method of selecting desired audio/data/visual information according to claim 1 further comprising the step of:

notifying the viewer that desired audio/data/visual information has been identified (126).

20. The method of selecting desired audio/data/visual information according to claim 19 wherein the viewer is notified while the viewer is interacting with audio/data/visual information signals.

21. The method of selecting desired audio/data/visual information according to claim 1 wherein the method is concurrently performed for a plurality of viewers.

22. A method of selecting desired audio/data/visual information comprising the steps of:

a) determining preferences of a user (100);

b) receiving a plurality of audio/data/visual signals (102);

c) comparing said preferences with the plurality of audio/data/visual signals to identify desired audio/data/visual information (122); and

d) providing access to the desired audio/data/visual information (124).

23. The method of selecting desired audio/data/visual information according to claim 22 further comprising the steps of:

selecting at least one of the plurality of audio/data/visual signals for comparison with said preferences (108).

24. The method of selecting desired audio/data/visual information according to claim 23 wherein the selection of at least one of the plurality of audio/data/visual signals is performed at least one of randomly, sequentially, and periodically.

25. An audio/data/visual signal selection system comprising:

an input device (14) for providing viewing preferences of a viewer; and

an information selector (22) for:

receiving a first plurality of audio/data/visual signals;

identifying, from the first plurality of audio/data/visual signals, a second plurality of audio/data/visual signals to be monitored wherein the second plurality of audio/data/visual signals is a subset of the first plurality of audio/data/visual signals;

comparing said viewing preferences with the second plurality of audio/data/visual signals to identify desired audio/data/visual information; and

providing access to the identified audio/data/visual information.

26. An audio/data/visual signal selection system comprising:

an input device (14) for providing preferences of a user; and

an information selector (22) for:

receiving a plurality of audio/data/visual signals;

comparing said preferences and the plurality of audio/data/visual signals to identify desired audio/data/visual information; and

providing access to the identified audio/data/visual information.

27. The audio/data/visual signal selection system according to claim 26 further comprising a memory (24), operatively coupled to the information selector, for storing at least a portion of said desired audio/data/visual information.

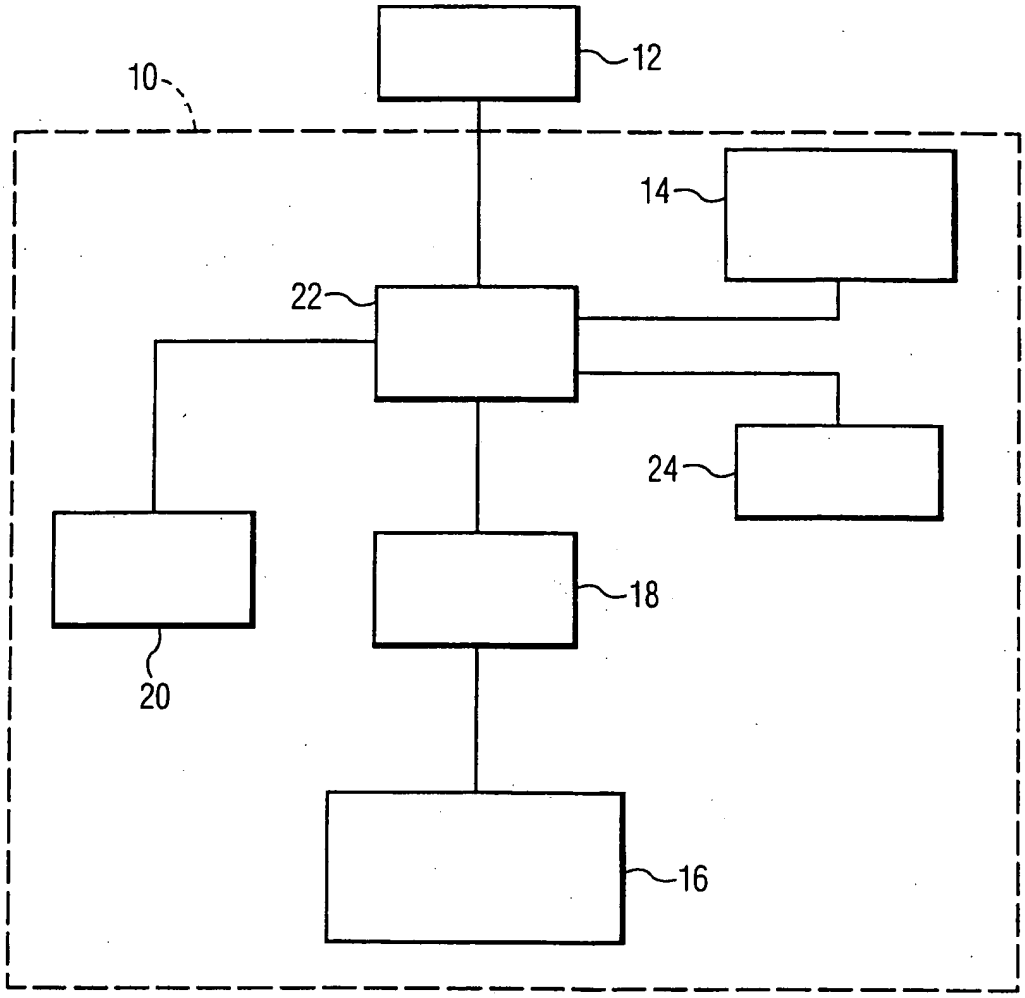

FIG. 1

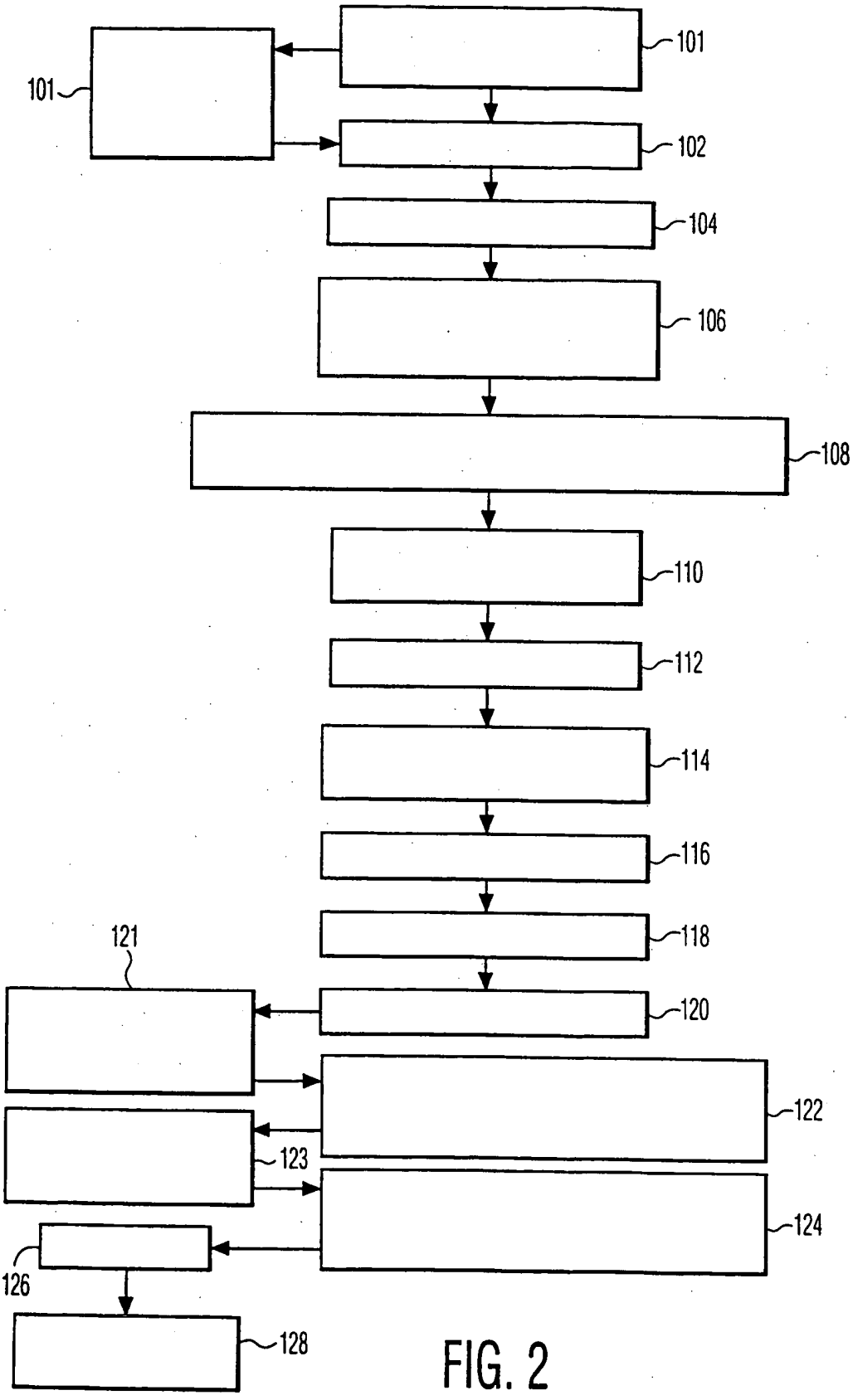
101
101
102
104
106
108
110
112
114
116
118
121
120
122
123
124
126
128

FIG. 2